Viele Unternehmen sammeln täglich riesige Mengen an Daten über ihre Kunden, Mitarbeiter, den Markt und viele andere Dinge, aber dieses Wissen ist einzeln betrachtet in bedeutungslosen Zahlen versteckt. Kombiniert man aber diese Daten könnten sie analysiert und visualisiert werden, um benutzerfreundliche und verständliche Diagramme zu erstellen, die helfen können, Marktveränderungen vorherzusagen oder den Gewinn durch intelligente Kostenreduzierung zu erhöhen.

Dies könnte entweder von einem Betreiber genutzt werden, der ansprechende Diagramme für seine Präsentationen und Berichte verwendet, oder von Datenanalysten, die diese Diagramme studieren und nützliche Informationen ableiten, um Fragen zu beantworten wie "Warum steigt der Markt in Großbritannien, obwohl es den Brexit gibt, und wie können wir das, was in dieser Region funktioniert, in andere Regionen bringen?

#Was sind vis-tools

Visualisierungs Applikationen sind eine Möglichkeit Daten als Grafiken widerzuspiegeln. Die Daten können auf vielfältige Art und weise zusammengestellt werden. Nicht jede Visualisierungs Applikation kann dieselben Datenquellen benutzen und verarbeiten. In Microsoft Word kann man zum Beispiel ausschließlich Microsoft Excel Tabellen für die Diagramme einfügen. In Spezialisierteren Applikationen können weitaus mehr Datenquellen verwendet werden wie Daten direkt aus einer Datenbank auslesen oder mit einer REST Schnittstelle von einem Server zu bekommen.

Unterschied JAvascript Zerocode

Die Entscheidung ob mit JavaScript Frameworks oder „Zero Code“ Frameworks gearbeitet werden soll muss bereits früh im Projekt geklärt werden weil es enorme Auswirkungen hat auf den Aufwand der eingeplant werden muss.

JavaScript Frameworks

JavaScript ist eine Skriptsprache und wurde erschaffen um CSS und HTML zu erweitern. Es kann verwendet werden um Websites Interaktiver zu machen sodass der Benutzer nicht jedes Mal eine vollständige neue Seite laden muss wenn er zum Beispiel eine neue Filteroption einer Liste auswählt. JavaScript kann aber noch viel mehr. Vgl <https://de.wikipedia.org/wiki/JavaScript#Verwendung>  
In unserem Fall könnte man JavaScript verwenden um interaktive Grafiken auf einer Website zu platzieren. Dies hätte den Vorteil das der Kunde das Diagramm nach seinen Wünschen anpassen könnte um die Daten anzuzeigen welche er gerade benötigt.

Nachteile bei der Verwendung von diversen JavaScript Frameworks sind zum Beispiel der Große Aufwand der von Nöten ist weil man einen vollständigen Server mit Datenbankzugriff selber ausimplementieren muss. Die Interaktivität muss man dabei ebenfalls selber integrieren.

Um einen besseren Überblick über die vorhandenen Libraries zu bekommen habe ich folgende Untersucht:

D3

„D3.js ist eine JavaScript-Bibliothek zur Manipulation von Dokumenten auf Basis von Daten. D3 hilft Ihnen, Daten mit Hilfe von HTML, SVG und CSS zum Leben zu erwecken. Die Betonung von D3 auf Webstandards gibt Ihnen die vollen Möglichkeiten moderner Browser, ohne sich an ein proprietäres Framework zu binden, und kombiniert leistungsstarke Visualisierungskomponenten mit einem datengesteuerten Ansatz zur DOM-Manipulation.“~cite{ bostock\_d3.js\_2019}

"Zero Code" Frameworks

Wie die Überschrift bereits vermuten lässt handelt es sich bei „Zero Code“ Frameworks um Applikationen die bereits vollständig implementiert worden sind. Der Arbeitsaufwand hält sich dementsprechend auch in gewissen Grenzen. Was aber nicht bedeutet das gar keine Zeit investiert werden muss denn zusätzlich zu dem Installations Vorgang muss auch eine Datenbank- oder eventuell REST-Schnittstelle implementiert werden. Nachdem unser Arbeitgeber MIC die Vorgabe gab das wir mit einem „Zero Code“ Framework arbeiten sollen folgt nun eine detailliertere Form der getesteten Frameworks:

Tableu:

Tableau ist ein kostenpflichtiges Datenvisualisierungstool, das für große Daten in Unternehmen optimiert ist. Obwohl es eine überwältigende Menge an Möglichkeiten hat und dies einige Zeit zum Erlernen benötigt, ist es besonders für Nicht-Techniker sehr einfach zu bedienen. Einmal richtig aufgesetzt können Datenfelder mit Hilfe von Drag n Drop einfach in das Diagramm gezogen werden und die Daten erscheinen im Diagramm. Es hilft auch bei der Auswahl der besten Grafik für jede Datenquelle indem es Vorschläge anzeigt. Hat der Benutzer ein Diagramm fertig kann er es auch in die Cloud stellen damit alle berechtigten Nutzer darauf zugreifen können.

Tablaeu Preise

Ein weiterer Wunsch von MIC war es das die verwendeten Technologien Gratis sind oder nicht viel kosten. Tableau erreichte diese Anforderung eindeutig nicht was die folgende Grafik über die Preispolitik zeigt.

Zusätzlich kann man noch extra Pakete dazu bestellen

Infogram ist ein Online-Werkzeug zur Erstellung von Grafiken und Diagrammen.

Es bietet 5 verschiedene Datenbanken (MySQL, Postgre, Amazon Redshift, Oracle und MS SQL-Server), aber keine REST-Schnittstelle, da es hauptsächlich dazu dient, hübsche Grafiken für Präsentationen und Berichte zu erstellen.

Außerdem kann man Daten per Upload importieren, mit Google Drive, Drop Box und JSON.

Infogram ist für seine Funktionen vergleichsweise teuer, ähnlich zu Tableau.

Nur eine kleine Auswahl an Funktionen und Grafiken

Chartblock

Chartblock ist ein sehr einfaches Werkzeug, mit dem man Diagramme in einer Wizard-Form(Schritt-für-Schritt Anleitung) erstellen kann. Zur Zeit der Entscheidung welches Tool verwendet werden soll hatte es keine Möglichkeit daten außer mit csv, Excel oder Manuell zu importieren, daher ist es für unseren Anwendungsfall derzeit nicht relevant

Dash

Dash ist eine Open-Source-Bibliothek für R und Python. Es wird kein JavaScript benötigt, aber Python. Mit Python schreibt man sowohl das Layout der Website sowohl als auch die Logik des Servers dahinter. Somit erreicht es einen ähnlich großen Funktionsumfang wie oben genannte JavaScript Frameworks mit dem bedeutenden Unterschied das Python mit wenigen Zeilen Code sehr viel mehr bewirken kann als JavaScript. Jedoch muss hier jedes Diagramm und Dashboard von einem Programmierer Entwickelt werden. Einfache Kunden können somit nur schwer Änderungen an dem Diagramm vornehmen oder gar neue erstellen.

Qlik

Qlik ist ein Desktop-Tool mit Schwerpunkt Data Analytics. Es kann aber auch Diagramme zeichnen.

Qlik kann Daten aus allen gängigen Quellen verarbeiten und es schlägt dabei auch vor wie Datensätze verbunden werden könnten was einiges an Aufwand spart. Auch bei der gemeinsamen Erstellung und Verteilung von Daten hilft Qlik. Sodass auch Nicht-Techniker die Informationen Nützen und verwenden können

PowerBI

PowerBI ist das Business-Intelligence-Werkzeug von Microsoft. Es ist einfach zu bedienen und hat sehr schöne Diagramme out-of-the-box. Leider konnte ich die REST-API nicht ausprobieren, weil man das online registrieren muss und Admin-Berechtigungen von MIC benötigt. Es gibt aber Testdaten mit denen man zumindest den Diagramm Teil der Applikation ausprobieren kann. Die Benutzung ist sehr einfach auch ohne technische Erfahrung und die Diagramme sind wirklich sehr schön

Kibana

Kibana ist ein grafisches Werkzeug, das auf Elasticsearch basiert. Elasticsearch ist eine Dokumentbasierte NoSQL Datenbank welche durch Features wie Volltextsuche und extrem schnelle Suchabfragen glänzt. Eine weitere tolle Eigenheit ist das Kibana sowohl als auch Elasticsearch schnell über Docker aufgesetzt werden können und sofort Startbereit sind und Daten über REST-Schnittstelle oder Zugriff auf externe Datenbanken importiert werden können. Kibana bietet aber auch out-of-the-box die Möglichkeit Musterdatensätze zu verwenden um einen ersten Eindruck zu erhalten. Kibana dient nicht nur zur Visualisierung, sondern auch zur Überwachung von Abfragen auf Daten. So kann z.B. kontrolliert werden von wem die Daten benutzt werden.

Vergleich der Visualisierungs Apps

Es gibt noch viel mehr Visualisierungsprogramme aber die wohl wichtigsten / bekanntesten sind hier gelistet. Eine paar der oben gelisteten Programme sind Nichenprodukte, es wird bestimmte Einsatzzwecke für Applikationen wie z.B. Chartblock geben jedoch werden diese nicht für die folgende Top 3 reichen.

Zur Erinnerung hier die Anforderungen von MIC für ein Visualisierungs Tool. REFERENZ EINFÜGEN

Driitter platz qlik

Qlik ist nicht schlecht, aber im Vergleich zu Kibana und PowerBI ist das Einrichten von Beziehungen und das Erstellen von Diagrammen etwas komplexer. Wenn man z.B. Daten von einer REST-Verbindung erhält, erzeugt es einige Spalten mehr als im Original-Datensatz, was sehr irritierend ist und für einen erhöhten Aufwand in der REST-Schnittstellenimplementierung sorgt. Ein weiterer wichtiger Punkt ist, dass Qlik-Benutzer sich über den hohen RAM-Bedarf beschweren welcher bei der geplanten Hochskalierung der Hardwareressourcen zu bedeutenden Kostenerhöhungen führen wird. Somit fällt Qlik trotz der guten Analytischen Funktionen auf Platz 3.

Zweiter Platz PowerBi

Die Entscheidung zwischen PowerBI und Kibana war eine schwere Wahl, weil PowerBI wirklich schön und einfach zu bedienen ist. Darüber hinaus hat es einige Features wie die "Voll Text ERKLÄRUNG REFERENZIEREN Abfragen" und "On-Click-Analyse", die sie sehr gut machen. Leider konnte ich die REST-API aufgrund eines komplexen Azure (Cloud von Microsoft) Registrierungsmechanismus nicht überprüfen. Ein entscheidender Faktor war die Auslieferung (Deployment) von PowerBI. Es gibt nur ein zusätzliches Tool namens "PowerBI Report Server", das nur dazu dient, Berichte innerhalb einer Firma zu verbreiten, die vorher in PowerBI Desktop gemacht wurden. Dieses eine Programm läuft auf einem Server in der Firma die es einsetzt. Der ganze Rest Verarbeitung der Visualisierung und der Analytischen Funktionen übernimmt die Microsoft Cloud Azure. Dadurch wird das Produkt nicht nur deutlich teurer, sondern die Firma wird auch sehr viel stärker an Microsoft gebunden.

Erster Platz Kibana

Kibana ist gleich in den meisten Punkten mit PowerBI, aber es gibt zwei wichtige Punkte, die Kibana ein wenig besser machen:

1. Kibana und Elasticsearch sind "out-of-the-box" für den Einsatz auf Servern konzipiert, was die Bereitstellung und den Zugriff von vielen verschiedenen Geräten aus erleichtert, ohne dass eine Desktop-Applikation installiert werden muss.

2. Besonders Elasticsearch ist (wie beworben) sehr schnell, da es für den Anwendungsfall gebaut ist für Analysefunktionen niedrige Zugriffszeiten zu bieten.

Technologien: Docker

Docker ist eine Open-Source Plattform für das Deployen von Softwareprodukten. Mit Docker kann man die verschiedenen Teile eines Softwareproduktes wie eine Datenbank, Server, Clients in einzelne Umgebungen verteilen sogenannte Container. Somit kann jede Komponente einzeln gestartet, konfiguriert, überwacht und getestet werden. Durch diese Modularität können Container fast unabhängig vom Betriebssystem arbeiten.

Container

Ein Container ist wie bereits in der Einleitung beschrieben ein abgeschlossenes Objekt welches ein oder mehrere Images (\ref{ssec:DockerImages}) besitzt und alle nötigen Dinge beinhaltet was die Software zum Laufen braucht wie die Runtime, Libraries, Einstellungen und natürlich der Code der Applikation. Wie in der Grafik (\ref {img:what-is-a-container}) beschrieben liegen die einzelnen Container auf der Docker Engine auf welche die Container verwaltet und die nötigen Ressourcen vom Betriebssystem zu Verfügung stellt. Dadurch wird sichergestellt das der Code so isoliert von der Umgebung ist wie möglich.

Images

Images sind extrem leichtgewichtig, weil sie auf das notwendigste reduziert sind und sicher da diese standardmäßig nicht von anderen Containern beeinflusst werden können. Ein gutes Beispiel dafür ist das Alpine Linux siehe Abbildung (\ref{img:docker-image-size}). Um dennoch mit anderen Containern und anderen Dingen wie dem File System interagieren zu können gibt es sogenannte Networks (\ref{ssec:DockerNetworks}).

Networks:

Networks verbinden die einzelnen Container miteinander. Es ist außerdem egal ob die Docker-Hosts unter Windows, Linux oder einer Mischung von beiden laufen. Die Networks verbinden beides Plattform unabhängig.